### **№18 ЕМТИХАН БИЛЕТ**

1. Нормализация. BatchNorm

2. Ассоциативті ережелер (немесе жөргегі бар сыра)

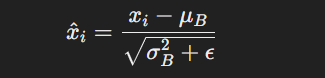
1. Нормализация: Пакетная норма (Batch Normalization)

Пакеттік қалыпқа келтіру-бұл нейрондық желілерді оқытудың жылдамдығын, өнімділігін және тұрақтылығын жақсарту үшін ұсынылған әдіс. Бұл әдіс ішкі ковариантты сдысу мәселесін азайтуға көмектесетін оқу деректерінің шағын пакеттері бойынша кез келген қабықтың шығуын қалыпқа келтіреді (изменения распределений входных данных каждого слоя в процессе обучения).

Механизм работы:

* Для каждой характеристики (или канала) в мини-пакете рассчитываются среднее значение и стандартное отклонение.
* Выходные данные нормализуются путём вычитания среднего значения и деления на стандартное отклонение для каждого признака, что обеспечивает приближение к стандартному нормальному распределению.

Нормализация, входные данные нормализуются:



𝜖ϵ — малое число, предотвращающее деление на ноль.

Масштабирование и сдвиг: Нормализованные данные затем масштабируются и смещаются с помощью обучаемых параметров 𝛾γ и 𝛽β:



- После нормализации применяются дополнительные параметры масштабирования и сдвига (\(\gamma\) и \(\beta\)), которые также обучаются в процессе тренировки сети.

Преимущества:

* Уменьшает зависимость от инициализации весов.
* Ускоряет обучение, позволяя использовать более высокие скорости обучения.
* Может служить некоторой формой регуляризации, слегка снижая переобучение.

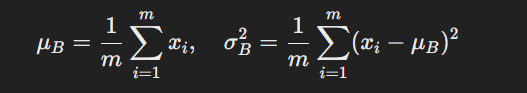
Основная концепция:

BatchNorm-тің негізгі идеясы-оқу кезінде әр қабаттың мини-пакеттік (batches) шығуын қалыпқа келтіру. Бұл алдыңғы қабаттардың параметрлеріндегі өзгерістерге байланысты оқу процесінде әр қабаттың кіріс үлестірімі өзгерген кезде пайда болатын ішкі ковариантты ығысуды (internal covariate shift) азайтады.

Процедура нормализации

BatchNorm исполняется следующим образом:

1.Вычисление среднего и дисперсии, для каждого признака в мини-пакете вычисляется среднее и дисперсия:



где 𝑥𝑖xi — это входные данные слоя, 𝑚m — размер мини-пакета.

**2. Ассоциативные правила (или пиво с подгузниками)**

Ассоциативті ережелер-бұл бөлшек сауда базалары сияқты транзакциялық деректердің үлкен жиынтығындағы элементтердің ортасында қызықты байланыстарды, жиі кездесетін заңдылықтарды, ассоциацияларды немесе корреляцияларды анықтауға арналған деректерді талдау әдістері.

образец "Пиво и подгузники":

Бұл сатып алушылардың сатып алуларын талдауда ұйымдастырылған бөлшек сауда аналитикасындағы танымал мысал. ертегіде дүкен деректерді талдау арқылы балаларына жөргек сатып алатын ер адамдар көбінесе сыра сатып алатынын анықтаған. Бұл қорытынды кросс-сатылымды ынталандыру үшін дүкендегі өнімдерді орналастыруды оңтайландыру үшін қолданылды.

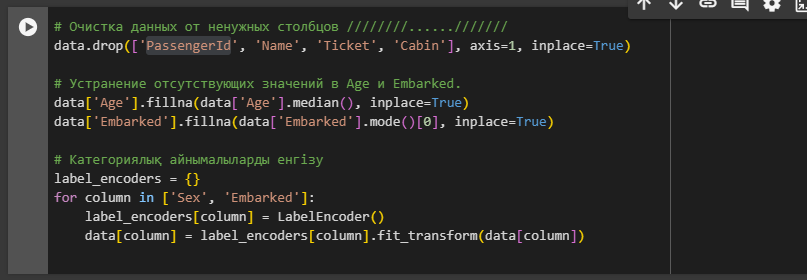
* Методы анализа:
* Поддержка (Support): Процент транзакций, хранящих товар A и B (например, пиво и подгузники).
* Уверенность (Confidence): Показывает, как часто элементы A и B возникают вместе, учитывая существование элемента A.
* Поднятие (Lift): Показывает, насколько чаще элементы A и B возникают вместе, чем можно было бы ожидать, если бы они были статистически независимы.

Ассоциативные правила обширно применяются в рекомендательных системах, управлении запасами, маркетинговых стратегиях и продвижении продуктов, предоставляя ценные инсайты для стратегического планирования и операционных решений.

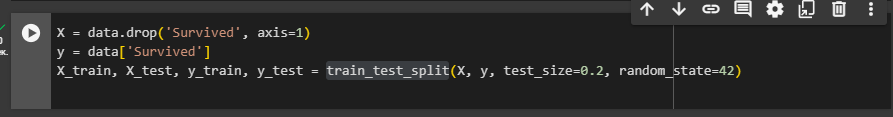
Есеп:

Google Colab: <https://colab.research.google.com/drive/1JKbZDdLW2ubYuwZ0b11R9D1Jp8m3z1CM?usp=sharing>

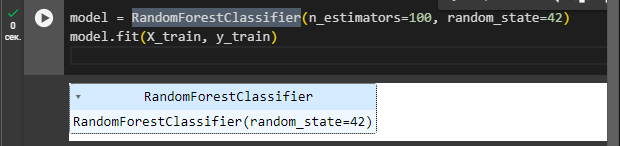
Деректерді өңдеу: кодта деректер қажет емес бағандардан тазартылады (мысалы, 'PassengerId', 'Name', 'Ticket', 'Cabin'), 'Age' және 'Embarked' бағандарындағы өткізіп алған мәндер толтырылады, ал 'Sex' және 'Embarked' категориялық айнымалылары жіктеу моделінде пайдалану үшін кодталады.



Деректерді оқыту және сынақ үлгілеріне бөлу: train\_test\_split функциясының көмегімен деректер оқу және сынақ жиынтықтарына бөлінеді, бұл модельдің өнімділігін бұрын көрмеген деректерде тексеруге мүмкіндік береді.



Жіктеу моделін оқыту: кездейсоқ орман моделі қолданылады (RandomForestClassifier), бұл көптеген белгілерді өңдеуге қабілетті және жалғыз шешім ағаштарымен салыстырғанда қайта оқытудан қорғалған күшті әдісі.



Модель өнімділігін бағалау: оқытудан кейін модель дәлдік көрсеткіштерін (accuracy\_score) және дәлдік, толықтық және F-өлшем сияқты көрсеткіштерді қамтитын егжей-тегжейлі жіктеу есебін (classification\_report) пайдалана отырып, сынақ деректерінде бағаланады.

